U.S. DEPARTMENT OF COMMERCE
NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION
NATIONAL WEATHER SERVICE
NATIONAL METEOROLOGICAL CENTER

OFFICE NOTE 258

A Very Simple Application of Kalman Filtering to Meteorological Data Assimilation

Norman A. Phillips
Development Division

JUNE 1982

ABSTRACT

The technique of Kalman filtering described by Ghil et. al. (1981) is applied to a 2-point model. This note is intended to make this technique more familiar at NMC and to point out its advantages and dangers. Two methods of allowing for ignorance of true dynamical factors are illustrated.

Contents

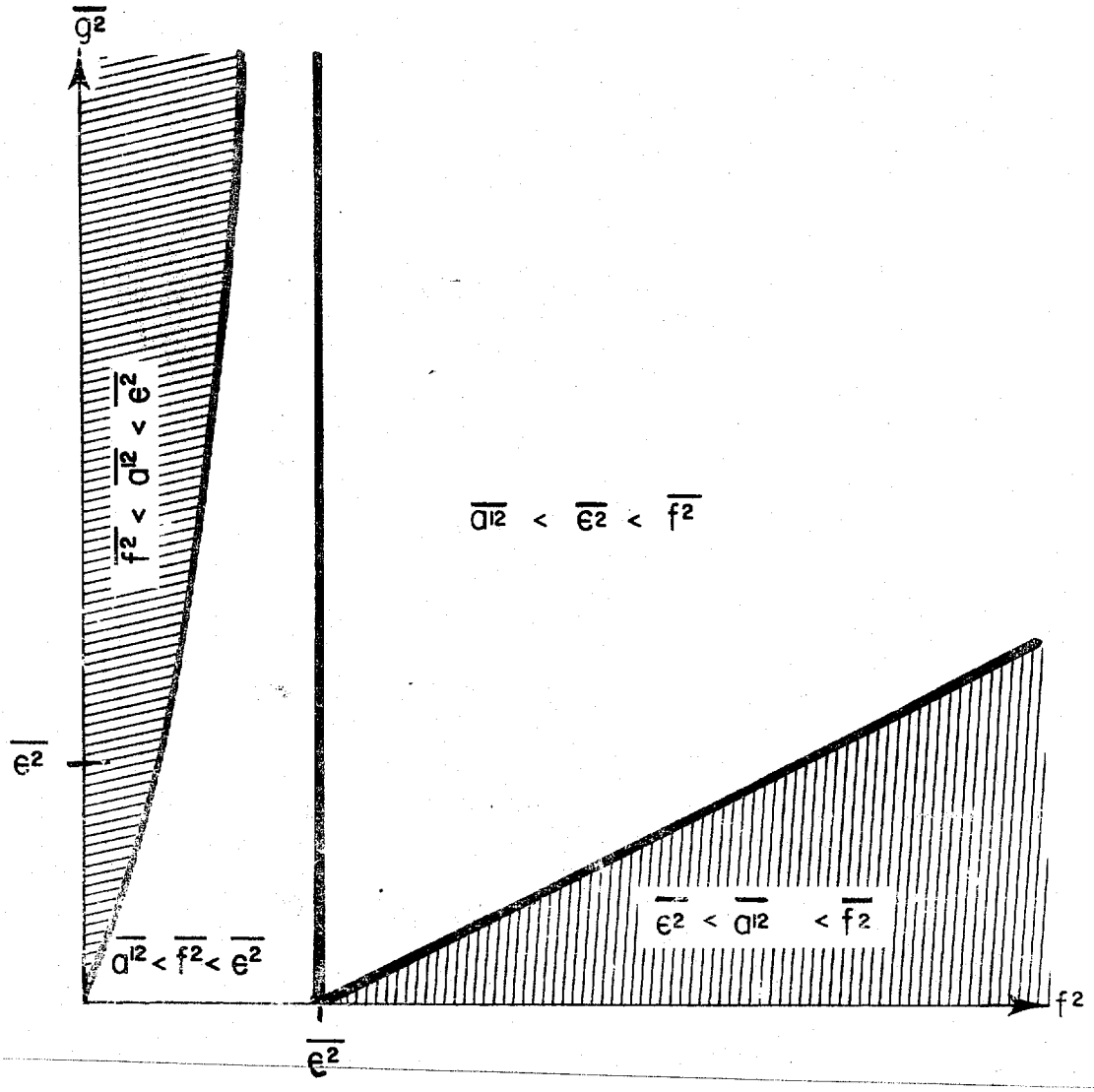

Figure 1. Analysis errors as a function of true ($\overline{f^2}$) and incorrect ($\overline{g^2}$) first guess errors.

1. Introduction

Ghil, et. al. (1981) have used a linearized one-dimensional version of the shallow water equations to explore the effect of Kalman filtering on a meteorological data assimilation system. This type of filtering is a method of statistically estimating the true state of a dynamical system whose evolution in time obeys a set of prediction equations and whose state is observed at intervals so the estimated state can be corrected (Gelb, et.al., 1974). The prediction equations deviate from the true system by random noise, and the observations are expected to have errors. The principal difference from a conventional meteorological data system lies in the attempt by the Kalman procedure to predict "first guess" error covariances accurately. Little thought has yet been addressed to this question in conventional meteorological systems. The paper by Ghil and collaborators therefore represents an important step in raising our awareness.

2. Effect of bad first guess error

Consider a single variable with a forecast value F and an observed value O. In the usual way, we form an analysed value A from these by a linear combination:

$$A = F + \alpha(O-F). \tag{1.1}$$

This will have an error

$$a = A-T = f + \alpha(\epsilon - f), \tag{1.2}$$

where f is the error in F and ϵ is the error in O. The expected value of a^2, denoted by $\overline{a^2}$, is given by $(1-\alpha)^2\overline{f^2} + \alpha^2\overline{\epsilon^2}$ if f and ϵ are uncorrelated. The value of α that minimizes this is

$$\alpha(opt) = \frac{\overline{f^2}}{\overline{f^2}+\overline{\epsilon^2}}, \tag{1.3}$$

and produces an analysis error

$$\overline{a^2} = \frac{\overline{\epsilon^2}\,\overline{f^2}}{\overline{\epsilon^2}+\overline{f^2}} < (\overline{\epsilon^2}, \overline{f^2})_{min} \tag{1.4}$$

Suppose however, that an incorrect value of $\overline{f^2}$, denoted by $\overline{g^2}$, is used. This would lead to a value α' :

$$\alpha' = \frac{\overline{g^2}}{\overline{g^2}+\overline{\epsilon^2}} . \tag{1.5}$$

But the true analysis error is given by

$$a' = f + \alpha'(\epsilon - f), \tag{1.6}$$

and we find that if f is the true error in F,

$$\overline{a'^2} = \frac{\overline{\epsilon^2}\left[\overline{\epsilon^2}\,\overline{f^2} + (\overline{fg})^2\right]}{(\overline{\epsilon^2}+\overline{g^2})^2} \geq \overline{a^2}. \tag{1.7}$$

Figure 1 delineates two regions of the $\overline{f^2}$, $\overline{g^2}$ plane where $\overline{a'^2}$ exceeds either the true first guess error $\overline{f^2}$, or the observation error $\overline{\epsilon^2}$. It is the second of these which is most disturbing, since one could do better in this region by omitting the prediction and analysis steps and simply using the observed value 0 for A. This occurs only when the (incorrect) first guess error $\overline{g^2}$ is less than the true first guess error $\overline{f^2}$.

In the remainder of this paper a simple model with two degrees of freedom will be used to illustrate the potential advantage of this approach and to also illustrate the potential dangers. In this note I will add a meteorological flavor by focusing attention on a dynamical system that may be unstable, but I will often shorten the discussion by omitting consideration of such matters as the requirement for an unbiased estimate. The book edited by A. Gelb (1974) also contains a discussion of

these aspects in the many illustrations it contains of simple mechanical and electrical systems.

3. The two point system and definitions

Let subscripts 1 and 2 denote two points at which analyses and observations are made. (In a more general treatment their location could vary with time.) We introduce the following notation (i,j = 1 or 2)

T_{in} = true state variable at point i, time n.
O_{in} = observation of T_{in}
ϵ_{in} = observation error = $O_{in} - T_{in}$
A_{in} = analysed value at point i, time n. (3.1)
a_{in} = analysis error = $A_{in} - T_{in}$
F_{in} = forecast value at point i, time n.
f_{in} = forecast error = $F_{in} - T_{in}$

An overbar will be used to denote an ensemble average. The ensemble we have in mind is a large collection of independent runs of an assimilation system, in which each run consists of repeated iterations on n of two successive steps:

$$A_{n-1} \text{—(prediction)}\rightarrow F_n$$
$$F_n + O_n \text{—(analysis)}\rightarrow A_n \qquad (3.2)$$

The ensemble average is taken, at fixed n, over all of the runs.

We will make several uses of the following relation. Let κ_1 and κ_2 represent any of the error quantities ϵ, a, or f. Suppose y and z are defined as fixed linear combinations of κ_1 and κ_2:

$$y = C_1 \cdot x_2 + C_2 \cdot x_2 + R_y,$$
$$z = C_3 \cdot x_1 + C_4 \cdot x_2 + R_z, \qquad (3.3)$$

where we have also added a random variable R ("noise"). If these equations are multiplied and then the statistical average performed, we

obtain

$$\overline{yz} = C_1 C_3 \overline{\nu_1 \nu_1} + (C_1 C_4 + C_2 C_3)\overline{\nu_1 \nu_2} + C_2 C_4 \overline{\nu_2 \nu_2} + \overline{R_y R_z} \quad (3.4)$$

if R is uncorrelated with ν_1 and ν_2. In other words, the linearity of (3.3) enables the covariance $\overline{yz}$ to be computed from the covariances $\overline{\nu_1 \nu_1}$, $\overline{\nu_1 \nu_2}$, $\overline{R_y R_z}$ and $\overline{\nu_2 \nu_2}$. This is a powerful result, but clearly depends on the linearity of (3.3) and the independence of the noise R. Its power results from the fact that while we cannot know the individual errors ν_1 and ν_2 in (3.3), we may be able to specify their covariances in (3.4).

4. The prediction system

In order to make use of (3.4) we need a linear prediction system. A simple one that is not devoid of meteorological meaning can be fashioned from a moving wave model:

$$\phi(x,t) = \text{const} \cdot e^{\lambda t} \cos(kx - \omega t)$$
$$= A(t)\cos kx + B(t)\sin kx \quad (4.1)$$

where

$$A(t) = \text{const} \cdot e^{\lambda t}\cos \omega t$$
$$B(t) = \text{const} \cdot e^{\lambda t}\sin \omega t \quad (4.2)$$

A and B satisfy the differential equations

$$dA/dt = \lambda A - \omega B,$$
$$dB/dt = \lambda B + \omega A. \quad (4.3)$$

If the assimilation times are denoted by $t = n\Delta t$, $n = 0,1,\ldots,$etc., the solution of (4.3) is

$$A_{n+1} = \nu A_n - \mu B_n,$$
$$B_{n+1} = \mu A_n + \nu B_n, \quad (4.4)$$

where

$$\nu = e^{\lambda \Delta t} \cos \omega \Delta t \,, \quad \mu = e^{\lambda \Delta t} \sin \omega \Delta t \,. \tag{4.5}$$

Alternatively, we can imagine (4.3) being solved by finite differences in time. If differences over n, n + 1 are used for the d/dt terms, and averages over n, n + 1 are used for the terms on the right side of (4.3)- i.e., an implicit forecast method - one arrives at the same time - stepping procedure as (4.4) except that ν and μ are replaced by

$$\begin{aligned} \nu &= [1 - (l^2 + w^2)] [(1-l)^2 + w^2]^{-1} \\ \mu &= 2w [(1-l)^2 + w^2]^{-1} \end{aligned} \tag{4.6}$$

in which l and w are defined by[1]

$$l = \lambda \Delta t / 2 \,, \quad w = \omega \Delta t / 2 \,. \tag{4.7}$$

Equations (4.4) are taken as the approximate equations of the true state, and their approximate character is assumed to be representable by simply adding random noise r to obtain the behavior of the true state T_1, T_2:

$$\begin{aligned} T_{1,n+1} &= \nu T_{1n} - \mu T_{2n} + r_1 \,, \\ T_{2,n+1} &= \mu T_{1n} + \nu T_{2n} + r_2 \,. \end{aligned} \tag{4.8}$$

[1] The computational stability is assured by the fact that $\sqrt{\nu^2 + \mu^2}$ approaches $1 + \lambda \Delta t$ as $\Delta t \to 0$.

The Ghil, et. al. approach is now to assume that the deterministic prediction system is used for the assimilation system, without the noise.

$$F_{1,n+1} = \nu A_{1n} - \mu A_{2n} ,$$
$$F_{2,n+1} = \mu A_{1n} + \nu A_{2n} . \qquad (4.9)$$

r is omitted from (4.9) because it is unknown in detail, so that (4.9) then gives the best individual estimate of $T_{i,n+1}$. Subtraction of (4.8) from (4.9) gives the error prediction system

$$f_{1,n+1} = \nu a_{1n} - \mu a_{2n} - r_1 ,$$
$$f_{2,n+1} = \mu a_{1n} + \nu a_{2n} - r_2 . \qquad (4.10)$$

Following the technique described at the end of section (3) now gives us a prediction formula for the first guess error covariance $(\overline{f_i f_j})_{n+1}$ as a function of the analysis error $(\overline{a_i a_j})_n$ at the previous assimilation time:

$$\overline{f_1 f_1}_{n+1} = \nu^2 \overline{a_1 a_1}_n + \mu^2 \overline{a_2 a_2}_n - 2\nu\mu \overline{a_1 a_2}_n + \overline{r_1 r_1}$$
$$\overline{f_2 f_2}_{n+1} = \mu^2 \overline{a_1 a_1}_n + \nu^2 \overline{a_2 a_2}_n + 2\nu\mu \overline{a_1 a_2}_n + \overline{r_2 r_2} \qquad (4.11)$$
$$\overline{f_1 f_2}_{n+1} = (\nu^2 - \mu^2) \overline{a_1 a_2}_n + \nu\mu (\overline{a_1 a_1}_n - \overline{a_2 a_2}_n) + \overline{r_1 r_2}$$

We have assumed no correlation between r_i and a_j.

Ghil, et.al. find it necessary, in their use of a linear shallow water system to introduce a Rossby-wave filtering step. This step can be omitted here because according to Phillips (1981), analyses are to be analyses of slow modes only, observations are to be used only after they have had fast modes subtracted, and the first guess error covariances are to refer to slow mode errors only. (4.8) and (4.9) therefore are to be interpreted as referring only to slow mode prediction. In fact all errors herein - observations, analyses and forecasts - refer only to

slow mode variables.

5. The analysis system

In the approach described by Ghil et. al., this is identical with the familiar optimum interpolation system used meteorologically, except that no approximation will be made in the first guess error covariances. To derive it in the present context, we first set up the equations:

$$A_1 = F_1 + \alpha(O_1 - F_1) + \beta(O_2 - F_2),$$
$$A_2 = F_2 + \gamma(O_1 - F_1) + \delta(O_2 - F_2). \tag{5.1}$$

All quantities are at a common time. These express the desired analysis as being equal to the first guess plus correction terms that depend on the differences between the observations and the first guess. By subtracting T_i from A_i, F_i, and O_i, we obtain the error relations

$$a_1 = f_1 + \alpha(\epsilon_1 - f_1) + \beta(\epsilon_2 - f_2), \tag{5.2}$$
$$a_2 = f_2 + \gamma(\epsilon_1 - f_1) + \delta(\epsilon_2 - f_2).$$

We square these and take the ensemble average as in (3.4). For simplicity, we assume that first guess error f_i and observation errors ϵ_i are not correlated. (This assumes that Δt is not too small in the prediction system or that ϵ_{im} is uncorrelated with ϵ_{im+1} .). We then choose α and β so as to minimize $\overline{a_1 a_1}$, and choose γ and δ to minimize $\overline{a_2 a_2}$. The result is

$$\alpha = [(\overline{\epsilon_2\epsilon_2} + \overline{f_2 f_2})\overline{f_1 f_1} - (\overline{\epsilon_1\epsilon_2} + \overline{f_1 f_2})\overline{f_1 f_2}] \div \Delta,$$
$$\beta = [\overline{\epsilon_1\epsilon_1}\,\overline{f_1 f_2} - \overline{\epsilon_1\epsilon_2}\,\overline{f_1 f_1}] \div \Delta$$
$$\gamma = [\overline{\epsilon_2\epsilon_2}\,\overline{f_1 f_2} - \overline{\epsilon_1\epsilon_2}\,\overline{f_2 f_2}] \div \Delta \tag{5.3}$$
$$\delta = [(\overline{\epsilon_1\epsilon_1} + \overline{f_1 f_1})\overline{f_2 f_2} - (\overline{\epsilon_1\epsilon_2} + \overline{f_1 f_2})\overline{f_1 f_2}] \div \Delta$$
$$\Delta = (\overline{\epsilon_1\epsilon_1} + \overline{f_1 f_1})(\overline{\epsilon_2\epsilon_2} + \overline{f_2 f_2}) - (\overline{\epsilon_1\epsilon_2} + \overline{f_1 f_2})^2$$

(The denominator Δ is positive.)

The resulting analysis error covariances can be evaluated by substituting these into the squared forms of (5.2). One finds (after some algebra)[2]

$$\begin{aligned} \overline{a_1 a_1} &= (1-\alpha)\overline{f_1 f_1} - \beta \overline{f_1 f_2}, \\ \overline{a_2 a_2} &= (1-\delta)\overline{f_2 f_2} - \gamma \overline{f_1 f_2}, \\ \overline{a_1 a_2} &= (1-\alpha)\overline{f_1 f_2} - \beta \overline{f_2 f_2} = (1-\delta)\overline{f_1 f_2} - \gamma \overline{f_1 f_1}. \end{aligned} \tag{5.4}$$

Our system is now complete, if we can assign values to

a) The prediction coefficients ν and μ.
b) The prediction noise: $\overline{r_1 r_1}$, $\overline{r_2 r_2}$ and $\overline{r_1 r_2}$.
c) The observation error: $\overline{\epsilon_1 \epsilon_1}$, $\overline{\epsilon_2 \epsilon_2}$ and $\overline{\epsilon_1 \epsilon_2}$.

That is to say, given a set of analysis errors $(\overline{a_i a_j})_n$, for the nth assimilation time, the prediction system (4.11) predicts the first guess errors $\overline{f_i f_j}_{n+1}$ at assimilation time n + 1. Then (5.4) - with the definitions

---

2 At NMC, the "diagonal" analysis errors $\overline{a_1 a_1}$ and $\overline{a_2 a_2}$ are calculated in the ITSOL subroutine. The value of $\overline{a_1 a_2}$ is ignored, however, since correlation between forecast errors and correlation between analysis errors are prescribed. It may be useful to relate the specific notation used here to the more general notation used in the paper by Ghil, et al. and the book by Gelb. Our analysis weights $\alpha, \beta, \gamma, \delta$ correspond to their "Kalman gain matrix" K. The first guess errors $\overline{f_i f_j}_n$ at assimilation step n correspond to their error covariance matrix $P_k(-)$ at step k. The analysis errors $\overline{a_i a_j}_n$ are equivalent to $P_k(+)$. In their treatment the vector of observation is given as a linear function (their matrix H) of the vector of true state variables (plus, the observation error vector). In our case H is an identity matrix. In a normal grid point model, however, H would be needed to interpolate between observation and grid point location.

(5.3) - enables the $\overline{a_i q_j}_{n+1}$ to be calculated for the n + 1 assimilation step.

This system is very stable when the noise r^2 is not zero. Steady states in which $\overline{f_i f_j}$ and $\overline{a_i a_j}$ no longer change from n to n + 1 are reached quickly in few cycles if $\overline{r^2} > 0$. An analysis of this is presented in the Appendix for the special case of $\overline{\epsilon_1 \epsilon_1} = \overline{\epsilon_2 \epsilon_2} = \epsilon^2$, $\overline{\epsilon_1 \epsilon_2} = 0$; with $\overline{r_1 r_1} = \overline{r_2 r_2} = r^2$, $\overline{r_1 r_2} = 0$. In this case the steady state values of $\overline{f_1 f_2}$ and $\overline{a_1 a_2}$ are zero, and $\overline{f_1 f_1} = \overline{f_2 f_2}$, $\overline{a_1 a_1} = \overline{a_2 a_2}$. For large r^2 with fixed ϵ^2, formulas (A.9) and (A.11) show that

$r^2 \to \infty$:

$$\overline{a_1 a_1} \to \epsilon^2 - \frac{\epsilon^4}{r^2} + O\left(\frac{\epsilon^6}{r^4}\right) \qquad (5.5)$$

The prediction system is here dominated by the large random forecast noise, but the Kalman method is still able to reduce the effect of this potentially large error source to a level where the resulting errors are somewhat less than the observational errors. The prediction system has been told the correct value of r^2, it must be remembered.

For $r^2 \to 0$ with fixed ϵ^2 the limits are more complex, and depend on σ.

a. $\sigma = \mu^2 + \nu^2 = 1$ (a neutral wave)

$$\overline{a_1 a_1} \to \epsilon r - \tfrac{1}{2} r^2 + O\left(\frac{r^3}{\epsilon}\right) \qquad (5.6)$$

(As shown in the Appendix, this state is approached very slowly.)

b. $\sigma = \mu^2+\nu^2 < 1$ (damped wave)

$$\overline{a_1 a_1} \to \frac{r^2}{1-\sigma}\left[1 - \frac{r^2}{\epsilon^2(1-\sigma)^2} + O\left(\frac{r^4}{\epsilon^4}\right)\right]. \qquad (5.7)$$

c. $\sigma = \mu^2+\nu^2 > 1$ (amplifying wave)

$$\overline{a_1 a_1} \to \epsilon^2\left(\frac{\sigma-1}{\sigma}\right) + \frac{r^2}{\sigma(\sigma-1)} + O\left(\frac{r^4}{\epsilon^2}\right). \qquad (5.8)$$

(These three cases agree with the equations (4.7c) and (4.7d) of Ghil, et.al., for the noise-free one-variable system.) In all cases $\overline{a_1 a_1}$ is less than the observational error $\overline{\epsilon^2}$, as required by (A.11). The value for $\overline{a_1 a_1}$ in the steady state is significantly less than $\overline{\epsilon^2}$ however only for the case of small r^2 and a wave that does not amplify too rapidly. Since (A.10) for the steady state can be written as[3]

$$\frac{1}{\overline{a_1 a_1}} = \frac{1}{\overline{\epsilon^2}} + \frac{1}{\overline{f_1 f_1}},$$

we see that this can only be achived by $\overline{f_1 f_1} < \overline{\epsilon^2}$, which requires small values of $\overline{r^2}$. (It will be seen in section 7 that this possibility is greatly limited when ν and μ are not known precisely.)

6. Comparison with NMC practice

Although the previous sections have considered only a 2-point system, it is possible to use it to expose the differences between the Kalman approach and the analysis methods at NMC. One obvious difference is that in (5.3)-(5.4) all observations are used in analysing each variable, whereas in practice only those close to the analysed point are used.

---

3 This is an example of the relation (4.2-19) in Gelb.

This is a limitation that is now becoming understood as a feature to be minimized, as far as computer power will allow, and we therefore ignore it here.

The computation of $\overline{a_1 a_2}$ in (5.4) is not a feature of the NMC system, however. It cannot be ignored in (5.4) because $\overline{a_1 a_{2n}}$ is needed in (4.11). But at NMC the equivalent of (4.11) can be characterized as

$$\begin{aligned} \overline{f_1 f_1}_{n+1} &= \overline{a_1 a_1}_n + \overline{G}^2, \\ \overline{f_2 f_2}_{n+1} &= \overline{a_2 a_2}_n + \overline{G}^2, \\ \overline{f_1 f_2}_{n+1} &= \rho \left[\overline{f_1 f_1} \cdot \overline{f_2 f_2} \right]^{1/2}_{n+1}. \end{aligned} \tag{6.1}$$

$\overline{G}^2$ represents an assumed growth of error during the time between up-date analyses. In this role it is analogous to the noise terms $\overline{r_1 r_1}$ and $\overline{r_2 r_2}$ in (4.11). ρ in (6.1) denotes a correlation coefficient computed from a reference set of forecasts verified over the United States. As such the forecasts on which it is based were in turn based on analyses whose errors reflect the excellent data from the continental rawinsonde network. We cannot expect this fixed ρ to reflect adequately the particular data mix present in other parts of the world and on every assimilation analysis n.

The potential importance of this correlation can be seen in the analysis coefficient equations (5.3). β and γ in those equations represent the analysis weight for observation 2 in the analysis at point 1 and for observation 1 in the analysis at point 2. The formula for β, for example, can be rewritten as

$$\beta = \frac{\overline{\epsilon_1 \epsilon_1}}{\Delta} \left\{ \sqrt{\frac{\overline{f_2 f_2}}{\overline{f_1 f_1}}}\, \rho^f - \sqrt{\frac{\overline{\epsilon_2 \epsilon_2}}{\overline{\epsilon_1 \epsilon_1}}}\, \rho^\epsilon \right\}, \tag{6.2}$$

where ρ^f and ρ^ϵ denote the correlation coefficients based on $\overline{f_1 f_2}$ and $\overline{\epsilon_1 \epsilon_2}$. It seems reasonable that there will often be a strong tendency for similarity in the ratio of first guess rms error at 2 points to the ratio of rms observing error at those two points. The size of β is then proportional to the differences between ρ^f and ρ^ϵ. The latter is prescribed (or at least assumed so!) for the observing system, whereas ρ^f may vary from one assimilation time to another as data mixes change.

The practical significance of the above argument for the importance of predicted as opposed to climatological first guess error correlations remains to be seen. The following section records an initial step toward considering this point.

7. Comparison with an approximate system

This can be done by constructing an approximate version of this simple two-point model patterned along NMC lines, and by evaluating the true error of the analyses made by this system.

With respect to the prediction equations, the approximate system will replace (4.11) by

$$\begin{aligned} \overline{f_1 f_1}_{n+1} &= \nu^2 \overline{a_1 a_1}_n + \overline{r_1 r_1}, \\ \overline{f_2 f_2}_{n+1} &= \nu^2 \overline{a_2 a_2}_n + \overline{r_2 r_2}, \\ \overline{f_1 f_2}_{n+1} &= \rho \left[\overline{f_1 f_1} \cdot \overline{f_2 f_2} \right]^{1/2}_{n+1}, \end{aligned} \tag{7.1}$$

where ρ is a precomputed correlation coefficient that is kept fixed during a "run". (7.11) ignores the μ effect in (4.4), as if the "wave" was not translating and forecast errors were a purely local process.

The approximate system will use the optimum interpolation equations (5.3)-(5.4), except that the last equation in (5.4) is irrelevant. However, the equations (5.4) only compute the apparent error of the approx-

imate system. It is more important for our purpose to know the true error of the approximate system. This can be done by running a parallel evaluation system. This will imitate the 12-hour forecast errors that are computed at NMC by comparing forecasts with verification data and will be similar to the computation of $\overline{a'^2}$ in section 2. This evaluation system will use the "correct" prediction equation (4.11) to compute the true first-guess error of the approximate system. The true analysis errors of the approximate system that are used on the right side of (4.11) must be derived by returning to (5.2). In this special use of (5.2):

A. $\alpha, \beta, \gamma, \delta$ are the coefficients computed by the approximate system according to (5.3), using the apparent first guess error covariances (7.1) of the approximate system.

B. $\overline{f_i f_j}$ is replaced by the true first guess error of the approximate system.

For example, the true $\overline{a_1 a_1}$ analysis error of the approximate system derived from (5.2) is given by

$$\overline{a_1 a_1} = \overline{f_1 f_1} + \alpha^2(\overline{\epsilon_1 \epsilon_1} + \overline{f_1 f_1}) + \beta^2(\overline{\epsilon_2 \epsilon_2} + \overline{f_2 f_2}) - 2\left[\alpha \overline{f_1 f_1} + \beta \overline{f_1 f_2} - \alpha\beta(\overline{\epsilon_1 \epsilon_2} + \overline{f_1 f_2})\right], \tag{7.2}$$

where α and β are computed according to A above, but $\overline{f_i f_j}$ has been computed from (4.11) using values of $\overline{a_i a_j}$ from (7.2). In this way the approximate system is associated with two sets of error statistics - the apparent set that it "thinks" are correct and the "true" set.

Numerical exploration of the iterative use of (4.11) and (5.3)-(5.4) shows that a stable stationary state is attained very quickly when $\alpha^2 > 0$ and that this state is independent of the $\overline{a_i a_j}$ (or $\overline{f_i f_j}$) that is assumed at n = 0. (The equilibrium state is derived mathematically

in the Appendix for a simple choice of $\overline{\epsilon_i \epsilon_j}$ and $\overline{r_i r_j}$.

The tests described in this section used the following choice of parameters.

$$\overline{\epsilon_1 \epsilon_1} = 2/3 \ , \ \overline{\epsilon_2 \epsilon_2} = 4/3 \ , \ \overline{\epsilon_1 \epsilon_2} = 0 .$$
$$\overline{r_1 r_1} = 1 \ , \ \overline{r_2 r_2} = 1 \ , \ \overline{r_1 r_2} = 0 .$$
$$\omega = 2\pi \div 3 \text{ days}, \qquad (7.3)$$
$$\lambda = \ln 2 \div 2.5 \text{ days},$$
$$\Delta t = 0.5 \text{ days}.$$

The finite-difference forms (4.6) were used, with the result that ν = 0.63231 and μ = 0.91833.[4]

The complete Kalman system under these conditions quickly arrives at the stationary state

$$\overline{a_1 a_1} = 0.489045$$
$$\overline{a_2 a_2} = 0.743905$$
$$\overline{a_1 a_2} = -0.027396 \qquad (7.4)$$
$$\overline{f_1 f_1} = 1.842071$$
$$\overline{f_2 f_2} = 1.690657$$
$$\rho_K = -0.079706$$

where $\rho_K = \overline{f_1 f_2} \, (\overline{f_1 f_1} \times \overline{f_2 f_2})^{-1/2}$.

The approximate system was run with the same choices of (7.3), as described earlier in this section, for five arbitrary preassigned values of ρ (-.6, -.3, 0, +.3, +.6) and a special value. This special value

---

[4] These correspond to use of the exact form with a period and doubling time of 3.25 and 3.18 days instead of the 3 days and 2.5 days given in (7.3).

was obtained by imitating an "off-line" empirical evaluation of ρ. A run using $\rho = 0$ quickly settled into a steady state in which ρ in the evaluation system was ρ_1. This value was then used as ρ in a second run of the approximate system and produced a new value ρ_2 in the evaluation system. ρ_2 was then used in the 3rd run of the approximate system. This process stabilized to a value of $\rho = -.077474$ after 5 substitutions. (A value of $-.074424$ was already obtained for ρ_1.) The closeness of these numbers to the true value in (7.4) suggests that an empirically assigned ρ is indeed a stable number if data type and location is fixed and if the dynamical system parameters are constant and known.

Figure 2 shows the apparent and true total analysis error $a_1\bar{a}_1 + a_2\bar{a}_2$ obtained for different ρ values, as well as the single Kalman point for reference. In this simple fixed observation system the empirically determined coefficient seems to be adequate. However, if the observing system was to change from day to day it could lead to significantly less accurate analyses than the ideal system would produce. An important aspect of the figure is the complete unreliability of the apparent analysis error as a guide to the correctness of the assumed ρ.

8. Behavior of an incorrectly known system.

Gelb, et. al. (1974, p 232) points out that incorrect specification of the dynamical constants (our μ and ν) can lead to malfunctioning of the Kalman system. In the meteorological context the linear prediction equations of section 4 would presumably be derived from linearization of the atmospheric equations of motion as suggested by Ghil, et. al. (1981). In this concept the "perturbation" is the error and our parameters μ and ν would be replaced by functions of the analysed field. Thus uncertainty will be introduced both by the linearization assumption that errors are

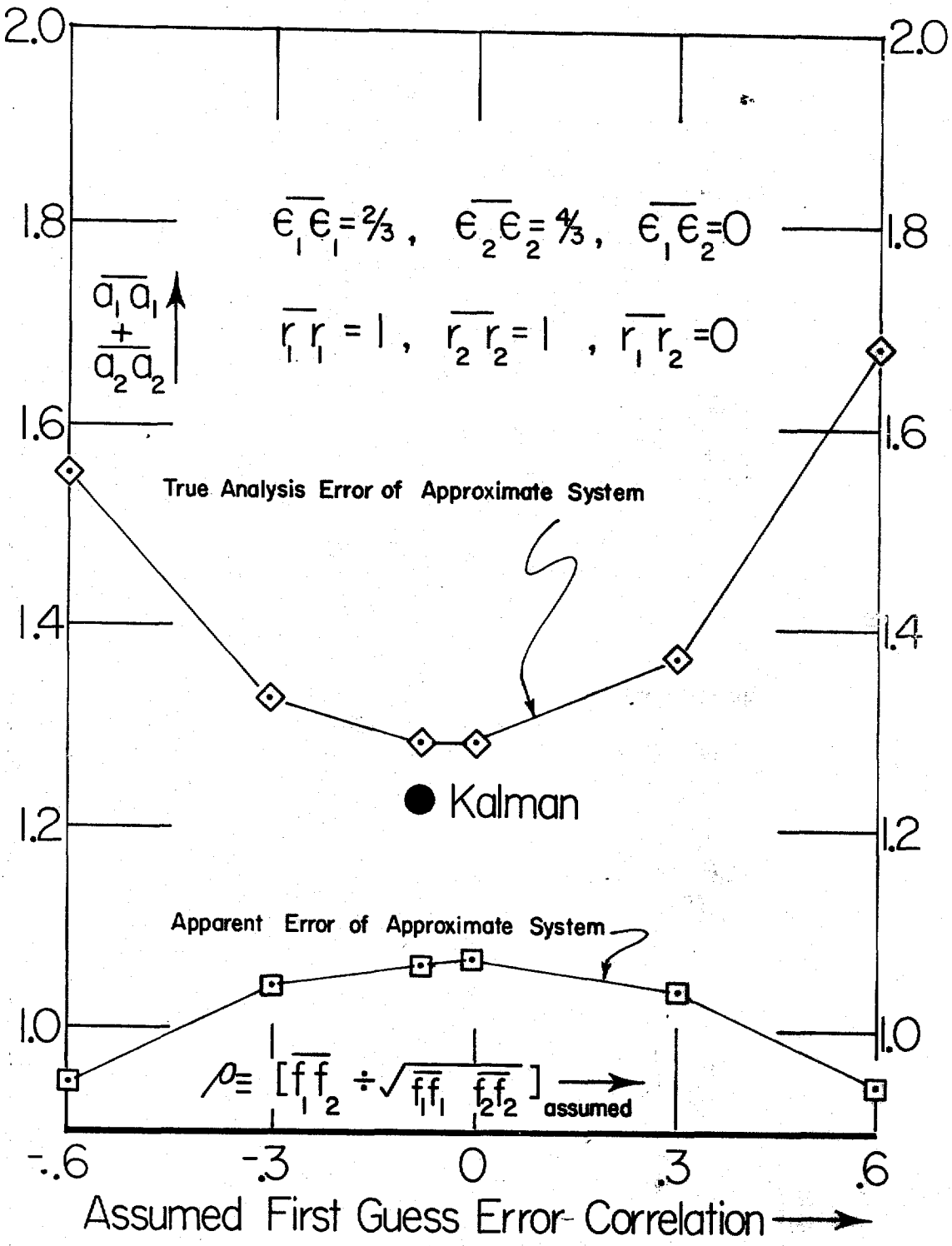

Figure 2. True and apparent analysis errors of the system based on empirical first guess error correlation ρ.

small and by uncertainty in the analysed fields. This problem can be illustrated to some extent with the present 2-point model. One possible corrective treatment, as one might expect, is to use a value for the system noise $\overline{n^2}$ that is large enough to accomodate, in some sense, our uncertainty about μ and ν. The problem is to use only as much added system noise as necesary, since very large noise leads to an analysis error that is not much less than the observational error [cf. (5.5)].

Suppose that the complete Kalman system consisting of (4.11) and (5.3)-(5.4) is used, but that incorrect values for some of the system parameters are used, as follows.

"Incorrect" System	Correct Value
$\overline{\epsilon_1\epsilon_1} = \overline{\epsilon_2\epsilon_2} \equiv \epsilon^2$,	$\overline{\epsilon_1\epsilon_1} = \overline{\epsilon_2\epsilon_2} = \epsilon^2$,
$\overline{\epsilon_1\epsilon_2} = 0$.	$\overline{\epsilon_1\epsilon_2} = 0$.
$\overline{n_1 n_1} = \overline{n_2 n_2} = \epsilon^2 R'^2$,	$\overline{n_1 n_1} = \overline{n_2 n_2} = \epsilon^2 R^2$,
$\overline{n_1 n_2} = 0$.	$\overline{n_1 n_2} = 0$.
ν', μ'	ν, μ
$\sigma' = \nu'^2 + \mu'^2$	$\sigma = \nu^2 + \mu^2$

(8.1)

In this example then, the observational errors are known correctly, but the only other system characteristics that are correct are the equality of $\overline{n_1 n_1}$ and $\overline{n_2 n_2}$, and the zero value for $\overline{n_1 n_2}$.

As shown in the Appendix, the incorrect system will eventually settle down to a steady state. In this state the apparent first guess errors are as follows:

$$\frac{\overline{f_1 f_1}}{\epsilon^2} = \frac{\overline{f_2 f_2}}{\epsilon^2} \equiv m = \frac{1}{2}\left[(R'^2+\sigma'-1)+\sqrt{(R'^2+\sigma'-1)^2+4R'^2}\right]$$

$$\overline{f_1 f_2} = 0. \tag{8.2}$$

The analysis weights $\alpha, \beta, \gamma, \delta$ in this state are given by (5.3).

Under conditions (8.2) they are

$$\alpha = \delta = \frac{m}{1+m}, \qquad \beta = \gamma = 0. \tag{8.3}$$

The true analysis errors of this system are then found by introducing (8.3) into (5.2). We obtain

$$\begin{aligned} \overline{a_1 a_1} &= (1-\alpha)^2 \overline{f_1 f_1} + \alpha^2 \epsilon_o^2 \\ \overline{a_2 a_2} &= (1-\alpha)^2 \overline{f_2 f_2} + \alpha^2 \epsilon_o^2 \\ \overline{a_1 a_2} &= (1-\alpha)^2 \overline{f_1 f_2}, \end{aligned} \tag{8.4}$$

where $\overline{f_i f_j}$ are now the true first guess errors of the incorrect system. These are in turn prescribed by the prediction system (4.11) using the correct ν, μ and $\overline{r_i r_j}$.

Let us define the following combinations of these true errors:

$$X \equiv (\overline{f_1 f_1} + \overline{f_2 f_2}) \div 2\epsilon^2, \qquad Z = \left[\frac{\overline{f_1 f_1} - \overline{f_2 f_2}}{\epsilon^2} \right]^2 + 4 \left[\frac{\overline{f_1 f_2}}{\epsilon^2} \right]^2. \tag{8.5}$$

Introduction of these into (8.4) and (4.11) leads to a simple system to describe their behavior after attainment of the steady state (8.2)-(8.3).

$$X_{n+1} = \Omega (X_n + m^2) + R^2, \tag{8.6}$$

$$Z_{n+1} = \Omega^2 Z_n \tag{8.7}$$

where

$$\Omega = \frac{\sigma}{(1+m)^2} \tag{8.8}$$

and m is defined by (8.2). (8.6) has the solution

$$X_n = \Omega^n X_o + (m^2 \Omega + R^2)(1 - \Omega^n) \div (1 - \Omega), \tag{8.9}$$

where X_0 is the value of X at $m = 0$ (the beginning of the steady state).

(8.4) shows that the analysis error being made by the incorrect system is

$$\frac{1}{2}(a_1\bar{a}_1 + a_2\bar{a}_2) = \epsilon^2\left[(1-\alpha)^2 X + \alpha^2\right] = \epsilon^2 \frac{X+m^2}{(1+m)^2} \equiv \epsilon^2 \bar{E} \tag{8.10}$$

Unbounded growth of the true first guess error X now produces unbounded growth of the analysis error because the observations are not being used correctly. [If $X = m$ - as it would ideally - (8.10) would equal $\epsilon^2 m/(1+m)$.] The incorrect system will therefore produce <u>disastrous</u> results unless $\Omega < 1$. This requires that

$$R'^2 + \sigma' + 1 + \sqrt{(R'^2 + \sigma' - 1)^2 + 4R'^2} > 2\sqrt{\sigma}. \tag{8.11}$$

If $R'^2 = 0$, and $\sigma' < 1$, this will be satisfied if and only if σ is also less than 1. If $R'^2 = 0$ and $\sigma' > 1$, it is satisfied if and only if $\sigma'^2 > \sigma$ (i.e., σ' need not be as large as σ.) We note however that (8.11) can always be satisfied by making R'^2 large enough.

A more meaningful criterion is to require that the averaged analysis error of (8.10) be less than ϵ^2, i.e., that $\bar{E} = (X+m^2)/(1+m)^2 < 1$. This criterion is relevant because ϵ^2 is the analysis error obtained by completely ignoring the assimilation system. (cf. the discussion in section 2.) Assuming for the moment that $\Omega < 1$, the asymptotic value of X from (8.9) is $(m^2\Omega + R^2)/(1-\Omega)$. Thus we require that

$$\bar{E} = \frac{X+m^2}{(1+m)^2} = \frac{R+m^2}{(1+m)^2 - \sigma} \tag{8.12}$$

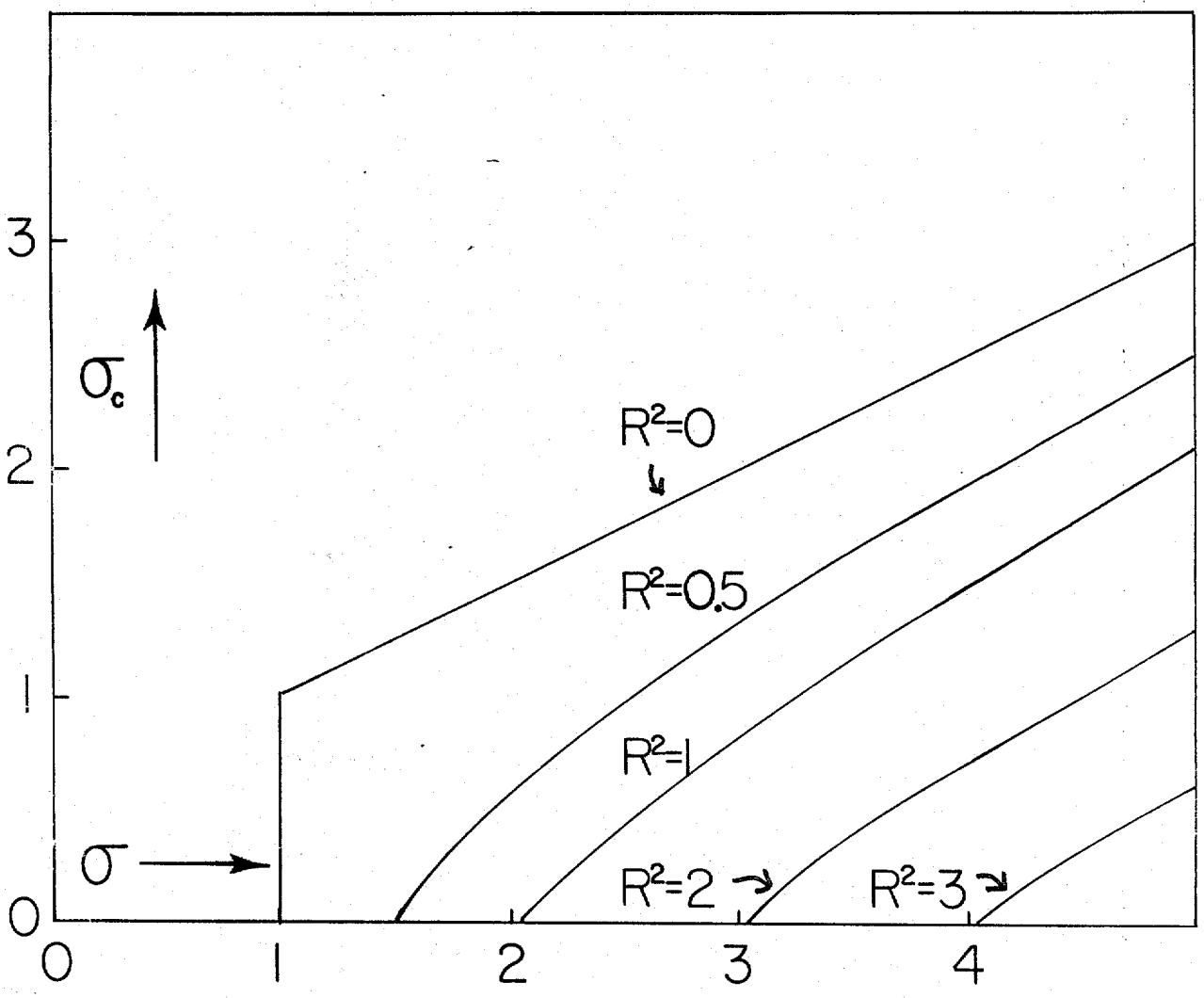

Figure 3. σ_c from (8.13) as a function of R^2 and σ .

be less than 1.

Consider first the case where the correct noise is used ($R'^2 = R^2$), but that σ' differs from σ. (8.12) and (8.2) can then be manipulated to show that E will be greater than 1 if

$$\sigma' < \sigma_c \equiv \frac{\sigma^2 - (R^2+1)^2}{2(R^2+\sigma-1)}, \quad \sigma > 1. \tag{8.3}$$

Figure 3 shows this relation. It contains the critical curves $\sigma' = \sigma_c$ in the σ, σ' plane for the indicated fixed values of R^2. If σ' is used with R^2, when the true values are σ and R^2, E will exceed 1 if σ' lies to the right of the particular R^2 curve shown in Figure 3. If σ' lies too far to the right the criterion $\Omega < 1$ will also be violated and the true analysis error will be unbounded.

The problem we face then is that if σ' underestimates σ we may have an assimilation system that is worse than no assimilation system at all. Figure 2 suggests that we might be able to avoid this by using an artificially increased value of R^2. The problem is to avoid going too far in this direction, since (5.5) shows that at large R^2 the <u>apparent</u> value of E will be

$$E_{app} \to 1 - \frac{e^2}{\lambda^2} = 1 - \frac{1}{R^2}. \tag{8.14}$$

The true value will be even larger and we would again be better advised to ignore assimilation.

Let us suppose that we can confidently set an upper limit σ_m to the values of σ' and σ.

$$\sigma \le \sigma_m, \quad \sigma' \le \sigma_m, \tag{8.15}$$

and explore the possibility of replacing the true system noise R^2 by

$$R'^2 = R^2 + (\sigma_m - \sigma'), \tag{8.16a}$$

or

$$\overline{n'^2} = \overline{n^2} + (\sigma_m - \sigma')\epsilon^2. \tag{8.16b}$$

This has the following intuitive justification.[5] ϵ^2 in (8.16b) can be interpreted as a conservative estimate of the analysis error. $(\sigma_m - \sigma')$, which is positive, represents the maximum possible underestimate of the true dynamical growth rate. The product therefore represents a reasonably bounded estimate of the additional error caused in the prediction step by our ignorance of the true σ. Returning then to (8.2) we find from (8.12) that the condition $E < 1$ becomes

$$R'^2 + \sqrt{(R'^2+\sigma'-1)^2+4R'^2} > R^2+\sigma-\sigma' \tag{8.17}$$

The definition (8.16a) for R'^2 satisfies this requirement even without the non-negative square root term. We also find that Ω is now given by

$$\Omega = \frac{4\sigma}{\left[R^2+\sigma_m+1+\sqrt{(R^2+\sigma_m-1)^2+4R'^2}\right]} \tag{8.18}$$

$$< \frac{4\sigma_m}{[R^2+\sigma_m+1]^2} < \frac{1}{1+R^2},$$

so that our use of the asymptotic value of X in (8.17) is justified.

Figures 4 and 5 illustrate the effect of using (8.16). In both figures $\sigma_m = 4$ and $R^2 = 1$. The solid curve in each of these figures gives the true normalized analysis error E for the

---

5 Gelb also discusses the addition of extra system noise to overcome errors in modeling system dynamics (p.279,ff).

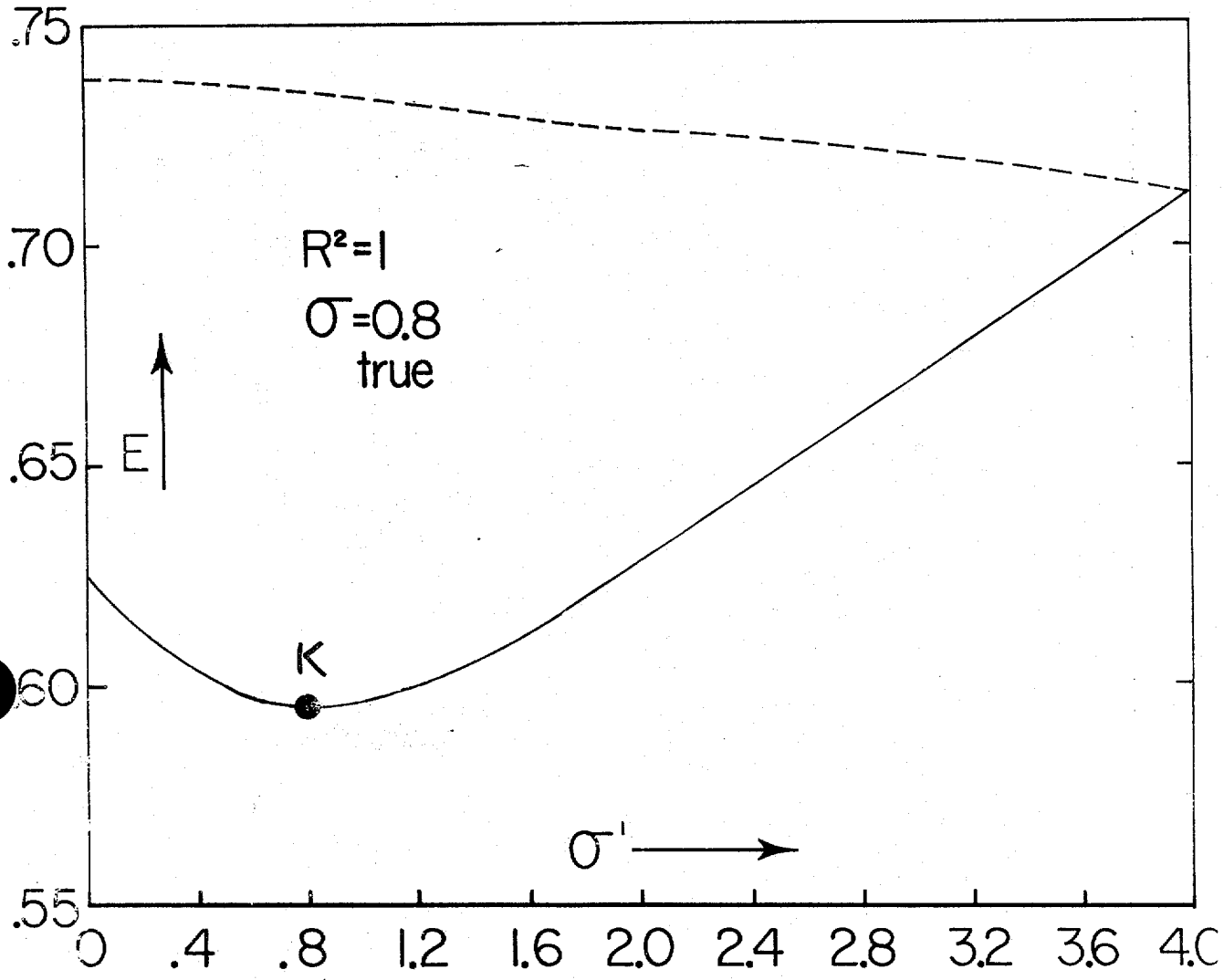

Figure 4. True normalized analysis error E as a function of the σ' used in the assimilation system when the true σ is 0.8 (solid curve). The dashed curve gives E when R^2 is replaced by (8.16a).

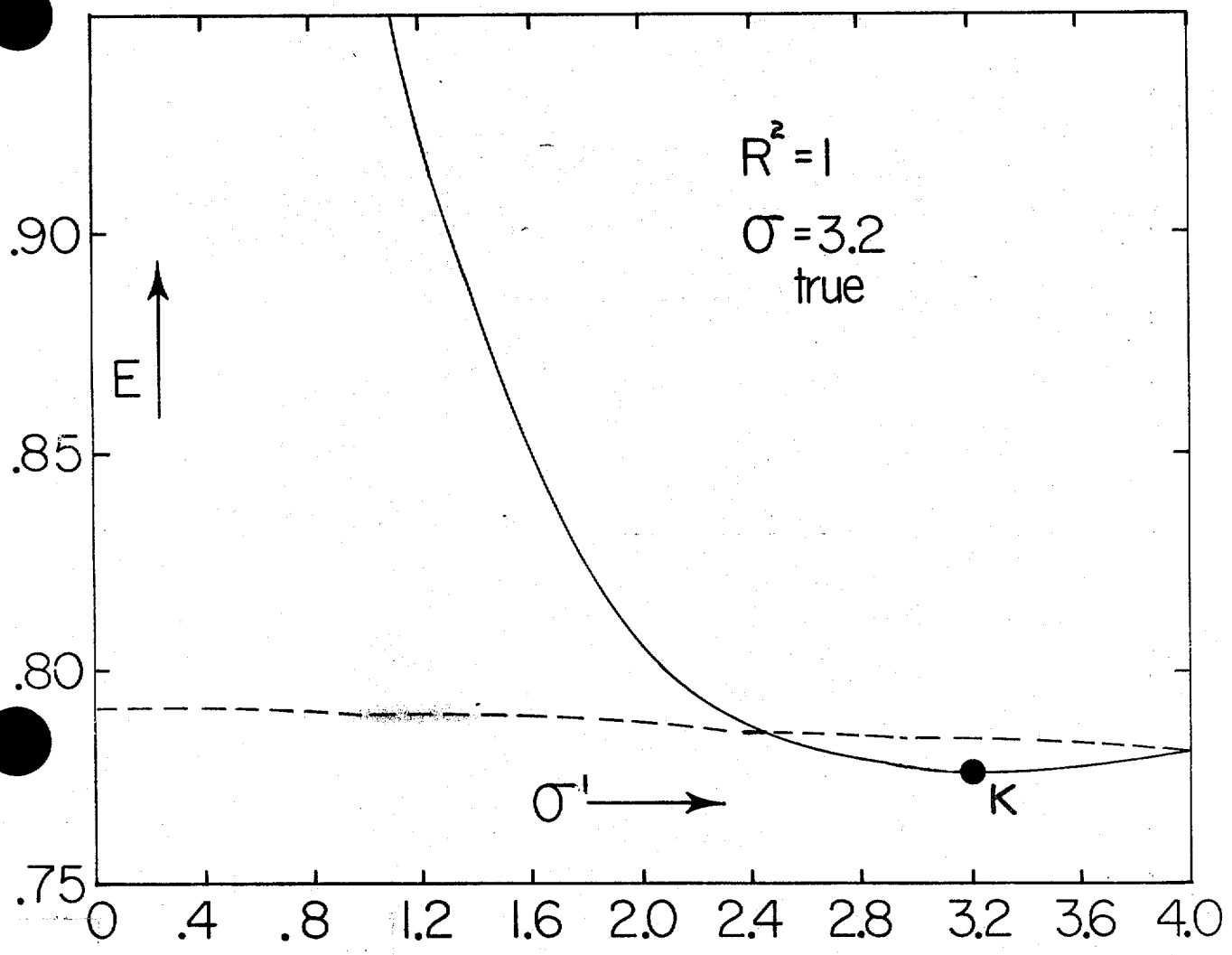

Figure 5. As for Figure 4 with σ (true) = 3.2.

"uncorrected" system that uses σ' (the abscissa) with the original $R^2=1$. Figure 4 has the true σ equal to 0.8 while Figure 5 has the true σ equal to 3.2. In Figure 4 the uncorrected system does not produce dire results even when it has $\sigma' = \sigma_m$. In Figure 5 on the other hand, the uncorrected system begins to do very poorly when $\sigma' < 1.6$ and produces $\bar{E}$ values greater than 1 when $\sigma' < .98$.

The dashed curves show in both figures the true normalized analysis error $\bar{E}$ that results from using σ' with R^2 replaced by (8.16a). They change very little with σ'. In Figure 4 the result is not as good as the uncorrected system, but $\bar{E}$ is still everywhere less than 1. In Figure 5, on the other hand, the loss in accuracy is trivial for $\sigma' > 2.4$, while the improvement in accuracy is marked for $\sigma' < 2.4$.

The corrective procedure (8.16) is therefore very beneficial when the true σ is greater than 1, but it produces somewhat worse results when the true σ is less than 1. Its overall benefit therefore depends on the likelihood that σ(true) exceeds 1 and the likelihood that when it does the estimate σ' of σ(true) will be significantly less than 1.

Other choices than (8.16) are possible in an attempt to cope with uncertainty in σ. For example, a surprisingly effective choice is to use a <u>fixed</u> value of σ such that $\bar{E}$ in (8.12) becomes equal to 1 when $\sigma = \sigma_m$.

$$\sigma' = \sigma_p \equiv \frac{\sigma_m^2 - (R^2+1)^2}{2(\sigma_m + R^2 - 1)} . \qquad (8.19)$$

			NOISE=	0.200, SIGP=	2.27500					
TRUE SIGMA	0.400	0.800	1.200	1.600	2.000	2.400	2.800	3.200	3.600	4.000
E(TRUE SIG)	0.225	0.310	0.409	0.501	0.575	0.632	0.677	0.713	0.742	0.766
E(FROM SIGP)	0.434	0.464	0.497	0.535	0.580	0.634	0.697	0.776	0.874	1.000
			NOISE=	0.400, SIGP=	2.06471					
TRUE SIGMA	0.400	0.800	1.200	1.600	2.000	2.400	2.800	3.200	3.600	4.000
E(TRUE SIG)	0.351	0.426	0.500	0.567	0.622	0.667	0.704	0.734	0.758	0.779
E(FROM SIGP)	0.478	0.507	0.541	0.579	0.622	0.673	0.733	0.805	0.892	1.000
			NOISE=	0.600, SIGP=	1.86667					
TRUE SIGMA	0.400	0.800	1.200	1.600	2.000	2.400	2.800	3.200	3.600	4.000
E(TRUE SIG)	0.437	0.501	0.560	0.613	0.657	0.694	0.725	0.750	0.772	0.790
E(FROM SIGP)	0.517	0.546	0.579	0.616	0.658	0.706	0.762	0.828	0.906	1.000
			NOISE=	0.800, SIGP=	1.67895					
TRUE SIGMA	0.400	0.800	1.200	1.600	2.000	2.400	2.800	3.200	3.600	4.000
E(TRUE SIG)	0.500	0.555	0.604	0.648	0.685	0.716	0.743	0.765	0.784	0.801
E(FROM SIGP)	0.551	0.580	0.612	0.648	0.688	0.734	0.787	0.847	0.917	1.000
			NOISE=	1.000, SIGP=	1.50000					
TRUE SIGMA	0.400	0.800	1.200	1.600	2.000	2.400	2.800	3.200	3.600	4.000
E(TRUE SIG)	0.550	0.597	0.639	0.676	0.708	0.735	0.758	0.778	0.795	0.810
E(FROM SIGP)	0.582	0.610	0.642	0.676	0.715	0.758	0.807	0.863	0.926	1.000
			NOISE=	2.000, SIGP=	0.70000					
TRUE SIGMA	0.400	0.800	1.200	1.600	2.000	2.400	2.800	3.200	3.600	4.000
E(TRUE SIG)	0.695	0.721	0.744	0.764	0.781	0.797	0.811	0.823	0.834	0.844
E(FROM SIGP)	0.697	0.721	0.747	0.775	0.805	0.838	0.874	0.912	0.954	1.000
			NOISE=	3.000, SIGP=	0.0					
TRUE SIGMA	0.400	0.800	1.200	1.600	2.000	2.400	2.800	3.200	3.600	4.000
E(TRUE SIG)	0.768	0.784	0.799	0.812	0.823	0.834	0.843	0.852	0.860	0.867
E(FROM SIGP)	0.770	0.790	0.811	0.834	0.858	0.883	0.910	0.938	0.968	1.000
			NOISE=	4.000, SIGP=	-0.64286					
TRUE SIGMA	0.400	0.800	1.200	1.600	2.000	2.400	2.800	3.200	3.600	4.000
E(TRUE SIG)	0.813	0.824	0.834	0.843	0.851	0.859	0.866	0.872	0.878	0.883
E(FROM SIGP)	0.819	0.836	0.854	0.872	0.891	0.911	0.932	0.954	0.976	1.000

Figure 6. True normalized analysis error E for different noise values R^2 and σ values for a system given the correct (R^2, σ), and for a system using R^2 and the σ_p defined by (8.19). σ_m is 4.0.

Figure 6 lists the results obtained when this is done, in comparison with the ideal results obtained by knowing the true σ . The results seem attractive even when the numerator in (8.19) is negative.[6]

Although both (8.16) and (8.19) have some promise as a way to react to uncertainty in σ , it must be remembered that these tests are not only for a 2-point model, they are performed under conditions where, because $\epsilon_1\bar{\epsilon}_2 = 0$ and $\lambda_1\bar{\lambda}_2 = 0$, the approximate and the true solution are such that $f_1\bar{f}_2$ and $q_1\bar{q}_2$ are zero. The steady state therefore depends only on $\sigma = \nu^2 + \mu^2$, and not on the relative size of ν and μ . Furthermore, the convenience of examining only the steady state results presumes that the incorrect σ and the true σ are fixed; day-to-day variations are not considered.

---

6 This treatment appears to be an example of the S_1 sensitivity measure described on page 245 of Gelb.

REFERENCES

Gelb, A. (Ed.), 1974 Applied optimal estimation. The M.I.T. Press, Cambridge, Mass. 374 pp.

Ghil, M., S. Cohn, J. Tavantzis, K. Bube, and E. Isaacson, 1981: Application of estimation theory to numerical weather prediction. In Dynamic Meteorology, Data assimilation methods. Springer-Verlag, New York, 330 pp.

Phillips, N., 1982: On the completeness of multi-variate optimum interpolation for large-scale meteorological analysis. To appear in Monthly Weather Review.

APPENDIX

The behavior of the system is easiest to see when the observational errors are equal ($\overline{\epsilon_1 \epsilon_1} = \overline{\epsilon_2 \epsilon_2} = \epsilon^2$) and uncorrelated ($\overline{\epsilon_1 \epsilon_2} = 0$), and the system noise errors are similar ($\overline{r_1 r_1} = \overline{r_2 r_2}$) and also uncorrelated ($\overline{r_1 r_2} = 0$). The results, in retrospect, appear to have some similarity with the one-point ("scalar") system analysed in section (4.3) of Ghil, et al.

We first define, for convenience,

$$m_1 = \frac{\overline{f_1 f_1}}{\epsilon^2}, \quad m_2 = \frac{\overline{f_2 f_2}}{\epsilon^2}, \quad m_3 = \frac{\overline{f_1 f_2}}{\epsilon^2}, \tag{A.1}$$

and

$$d_1 = \frac{\overline{a_1 a_1}}{\epsilon^2}, \quad d_2 = \frac{\overline{a_2 a_2}}{\epsilon^2}, \quad d_3 = \frac{\overline{a_1 a_2}}{\epsilon^2} \tag{A.2}$$

$$\overline{r_1 r_1} = \overline{r_2 r_2} \equiv R^2 \epsilon^2 .$$

The O/I analysis equations reduce to

$$\begin{aligned} d_1 &= [m_1 + (m_1 m_2 - m_3^2)] \div \psi, \\ d_2 &= [m_2 + (m_1 m_2 - m_3^2)] \div \psi, \\ d_3 &= [m_3] \div \psi, \\ \psi &= (1+m_1)(1+m_2) - m_3^2 . \end{aligned} \tag{A.3}$$

The prediction equations are

$$\begin{aligned} m_1' &= \nu^2 d_1 + \mu^2 d_2 - 2\nu\mu d_3 + R_1^2 \\ m_2' &= \mu^2 d_1 + \nu^2 d_2 + 2\nu\mu d_3 + R_2^2 \\ m_3' &= \nu\mu d_1 - \nu\mu d_2 + (\nu^2 - \mu^2) d_3 , \end{aligned} \tag{A.4}$$

where the prime denotes m_j for the step after the m_j in (A.3). It is now convenient to define

$$\begin{aligned} \alpha &= m_1 + m_2 , \quad \beta = (m_1 - m_2)^2 + 4 m_3^2 , \\ \sigma &= \nu^2 + \mu^2 . \end{aligned} \tag{A.5}$$

The system reduces to one in the two variables $\varkappa$ and ζ :

$$\varkappa' = \left(\frac{\sigma}{\psi}\right)\left[\varkappa + \tfrac{1}{2}(\varkappa^2 - \zeta)\right] + 2R^2, \tag{A.6}$$

$$\zeta' = \left(\frac{\sigma}{\psi}\right)^2 \zeta, \tag{A.7}$$

where in the new notation

$$\psi = 1 + \varkappa + \tfrac{1}{4}(\varkappa^2 - \zeta).$$

A stationary state for non-zero ζ exists in (A.7) if $\psi = \sigma$. But this is not possible for $\varkappa$ in (A.6), since $\varkappa^2 - \zeta = 4(m_1 m_2 - m_3^2)$ is ≥ 0, as is R^2 . Therefore ζ must be zero in a steady state. From this we conclude that

Steady state:

$$m_1 = m_2 \equiv m_a, \tag{A.8}$$

$$m_3 = 0$$

(A.6), with $\varkappa' = \varkappa = 2m_a$ now leads to the solution

$$m_a = \tfrac{1}{2}\left[(R^2 + \sigma - 1) + \sqrt{(R^2 + \sigma - 1)^2 + 4R^2}\,\right]. \tag{A.9}$$

The analysis errors in the steady state are also equal and uncorrelated

$$\overline{a_1 a_1} = \overline{a_2 a_2} = \epsilon^2 \frac{m_a}{1 + m_a} \tag{A.10}$$

The convergence to this steady state in its vicinity depends on smallness of the ratio

$$\frac{\sigma}{\psi_a} = \frac{\sigma}{(1 + m_a)^2}. \tag{A.11}$$

This is less than 1 for $\sigma<1$ (a damped wave) because $m_2 \geq 0$. For $\sigma>1$ we note from (A.9) that $1+m_2=\sigma$ when $R^2=0$, giving us $\sigma/\psi_2(R^2=0)=\frac{1}{\sigma}<1$. m_2 increases with R^2 so that σ/ψ_2 is less than 1 for all R^2 when $\sigma>1$. However, a special case occurs for $R^2=0$ and $\sigma=1$. This has $m_2=0$, but $\sigma/\psi_2=1$. In this case the stationary state has zero errors but convergence to it will be extremely slow or non-existent.

The convergence problem at $R^2=0$ for $\sigma=1$ is made clearer by noting that

$$\frac{\sigma}{(1+m_2)^2} = \sigma\left[\frac{\epsilon^2}{\epsilon^2+(\overline{f_1 f_1})_2}\right]^2 \qquad \text{(A.12)}$$

and

$$(\overline{f_1 f_1})_2 = \frac{1}{2}\left\{\left[r^2+\epsilon^2(\sigma-1)\right]+\sqrt{\left[r^2+\epsilon^2(\sigma-1)\right]^2+4r^2\epsilon^2}\right\} \qquad \text{(A.13)}$$

For $\sigma=1$ and $\epsilon \gg r/2$, $(\overline{f_1 f_1})_2 \sim r\epsilon$ and

$$\frac{\sigma}{(1+m_2)^2} \to \left(1+\frac{r}{\epsilon}\right)^{-2} \qquad \text{(A.14)}$$

The slow convergence that this implies for $r \to 0$ is therefore easier to understand as a slow convergence arising from very inaccurate observations (large ϵ). Evidently when $\sigma=1$ in this simple model we cannot distinguish between these limits because the term in (A.13) of highest order in ϵ is multiplied by $\sigma-1$. The analysis accuracy $\overline{a_1 a_1} \to \epsilon r$ in this case is of course sensitive to whether ϵ is large or r is small.